DOCUMENT RESUME

ED 044 245 RE 002 946

AUTHOR	Leong, Che Kan
TITLE	An Experimental Study of the Vocabulary of Written Chinese among Primary III Children in Hong Kong.
PUB DATE	8 May 70
NOTE	8p.; Paper presented at the conference of the International Reading Association, Anaheim, Cal., May 6-9, 1970
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.50
DESCRIPTORS	Auditory Perception, Aural Learning, *Aural Stimuli, *Basic Vocabulary, *Chinese, *Grade 3, *Orthographic Symbols, Tape Recordings

ABSTRACT

The depth and range of the corpus of written Chinese symbols used by 9-year-old Hong Kong children were studied in order to compile a minimum written Chinese vocabulary list. Error patterns of children were also analyzed. After presenting some characteristics of the language and reviewing previous studies, the methodology was described. The sample size ranged from 407 to 461 children who were matched for age, sex, and academic performance. The 1,851 different symbols or characters compiled from survey returns were incorporated into 30 prose passages of 100 characters each (with about 40 percent recurring characters) and recorded on tape. Taped explanations of the passage content preceded three repetitions of the passage during which the children listened, transcribed, and reviewed. From the accumulation of correct and incorrect responses, a facility index was derived for each of the 1,851 characters. It was concluded (1) that 719 characters were reproduced by at least 70 percent of the children and should be regarded as the minimum core of written vocabulary at this level and (2) that the mean number of strokes for characters of median difficulty was 14. Further research in auditory reception of language and encoding of the symbols received is recommended. References are included. (CL)

ED044245

Che Kan Leong
Assistant Professor of
Education,
Special Education Department,
University of Saskatchewan,
Saskatoon, Saskatchewan,
Canada.

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

INTERNATIONAL READING ASSOCIATION
FIFTEENTH ANNUAL CONVENTION - MAY, 1970
Reading at the International Level - Friday, 8th May, 1970 at 4.00 p.m.

"An Experimental Study of the Vocabulary of Written Chinese among Primary III Children in Hong Kong"

RE002 946

This paper reports the results of the first stage of a longitudinal project which aims at studying the minimum vocabulary of written Chinese of primary school children in Hong Kong. The knowledge of this aspect of the language calls for the active ability to reproduce written symbols as contrasted with the more passive command of the reading vocabulary. In this study the main aim was to investigate both the range and depth of thr corpus of written Chinese symbols or characters used by children in the third year of primary schooling with a view to compiling a list of minimum written Chinese vocabulary. A subsidiary aim was to analyse error patterns of the children for more effective teaching and for the construction of diagnostic tests.

At the outset it was decided that the materials collected must be based on the language repertoire of the children and that teachers would be actively involved in seminars, field-work, marking of scripts and feed-back of information to pupils so as to maximise the effects of the experiment.

II Some Characteristics of the Chinese Language

Generally, the Chinese language can be classified as monosyllabic, islating and analytic.

Chinese can be said to be monosyllabic as each character is a phonemic element or a syllable. But this does not apply to words which may contain more than one phonemic element. An example is the equivalent of the word "madam" (夫人) where each character by itself means "husband" (夫) and "person" (人). Chinese does not have, as certain other languages, disyllabic or polysyllabic stem-words.

Chinese can be said to be isolating in contrast to inflexional languages where there is a merging of semantically distinct features either in a single bound form or in closely united bound forms or to agglutinative languages where bound forms follow one another. The character (人) means equally "man" "man's", "men", "men's" "the man", "the man's", "the men" and "the men's". The use of classifiers surmounts the difficulty.

Chinese is better described as analytical where each word is a one-syllable morpheme or a compound word or phrase-word and where there are few bound forms. This is in contrast to synthetical languages which use many bound forms.

In addition, Chinese is distinguished by the use of pitch as primary phoneme. Northern Chinese differentiates four of these: a) high level,

e.g. "mother" (媽) pronounced as /ma/; b) high rising - "hemp" (蘇) as /ma/; c) low rising - "horse" (馬) as /ma/ and d) low falling - "scold" (罵) as /ma/. Southern Chinese further adds two tones, viz., the high and the low tones for each of the above pitch levels and an intermediate tone for the low falling pitch as well, making a total of nine tones. The bases of all these tonal differences are: a) duration - the relative length of time through which the vocal organs are kept in a position and b) stress - intensity or loudness.

From the foregoing it can be seen that there is a large number of homonyms and homophones. Thus the character with the short vowel sound "i" can mean something like thirty-eight different words.

III. Review of Previous Studies

Very broadly, vocabulary studies fall into three categories. One group of researchers such as Burroughs (1957) studies the oral vocabulary of children. The second group such as Rinsland (1947) studies the work written by children. The third group studies children's works, Thorndike's word count (1944) being a classic example.

In the case of the Chinese language Chuang (1938) compiled a "fundamental vocabulary" of 5262 characters. The list is divided into characters most frequently used with about 2800 characters; those less frequently used and those rarely used with about 1200 characters in each of these latter two categories. More recently, Kennedy (1954) edited the minimum vocabularies of written Chinese. The list of 1020 characters includes those which are used at Yale University for checking purposes. The list of 2421 characters is the official list for adult education in mainland China. This is further broken down into: a) 1010 most frequently used

characters forming with another 490 characters a limiting list of 1500 characters; b) a supplementary list of 500 characters and c) a list of technical terms of 421 characters.

While the above two studies are useful they would not be entirely applicable to the present study. Chuang's list is over thirty years old and may not accord with contemporary usage. Kennedy's list is helpful but the selection of characters needs to be refined.

IV. Methodology of Present Study

1) Basal Year of Study

Children in the third year of primary schooling (age 9+) were selected for the study as by this time they will have mastered the mechanics of reading and writing and as there is evidence that the third year in the primary is a critical period for these skills at the higher level. Moreover, results from the annual achievement testing programme in the basic subjects of languages and primary mathematics were available from this grade upwards for comparison and further analysis.

2) Selection of Test Materials

Previous studies show that by the end of six years of primary schooling, children will have acquired a Chinese reading vocabulary of about 3500 to 4000 characters. At the end of the third year of primary schooling children will have learnt about 1500 to 2000 characters.

In the selection of test materials care was taken to ensure that: a) the characters be serviceable and within the experiential background of the children and b) the written vocabulary be contained in the reading vocabulary.

Altogether 1851 different or token characters were sampled. These were incorporated into thirty connected prose passages each dealing with a specific topic such as "A Picnic," "Santa Claus" and each containing

about 100 characters. Of the corpus of characters tested recurring ones amounted to about 40% and for each passage about 60 new or different characters were tested.

3) Sampling

The experiment was conducted from November, 1966 to May, 1967. A cluster sample of three to four Government primary schools was used for six to eight passages. The sample size ranged from 407 to 461 and was in excess of the 390 required for a finite population of 15000 Government Primary III children with an assurance of an error not exceeding 1% at the 95% confidence level. There was no significant difference in respect of age, sex and academic performance between sample and population as shown by chi-square tests.

4) Experimental Procedure

a) The "language perception test" technique through the use of the tape recorder was employed. Each passage of thirty minutes duration including explanations was recorded on tape. This made possible the administration of the taped "tests" in a standardised situation by the class teacher.

b) Before the test proper, the content of each passage was explained fully and visual aids was used where possible. After the recorded explanation the passage was reproduced from the tape recorder three times at dictation speed. The children listened on the first occasion, wrote down each character on the second occasion and reviewed their writing on the third.

c) After the "dictation" the children's versions were marked by the teacher. Each character was endorsed as either correct or incorrect.

Gradations and patterns of errors were ignored at this stage. The manual marking enabled teachers to feed back relevent information to the pupils.

d) Analysis made possible the derivation of a facility index for each character. The facility index is expressed as the percentage of all participating children who wrote a character correctly. The 40% or so recurring characters provided a reliability estimate for these characters. Table I below shows the number of characters grouped by facility levels:

Table I

Facility Levels (F)	No. of Characters	
90%-100%	292	
80%-89%	229	
70%-79%	198	total 719
60%-69%	169	
50%-59%	183	total 352
40%-49%	160	
30%-39%	176	
20%-29%	184	
10%-19%	137	
0%-9%	123	total 780
		total 1851

V. Discussion of Results

1) The 1851 token characters are grouped under type characters in "dictionary" order with radicals or lexical units followed by characters arranged according to the complexity or numbers of strokes within each character. A facility index is given for each character as shown in Appendix I.

2) The arbitrary threshold score of 70% facility index is taken as the lower limit to indicate those characters which are reproduced correctly by nearly three quarters of the Primary III children. Thus the 719 characters in Table I above should be regarded as the minimum core of written

vocabulary of children at this level. The value of 70% ensures the possibility of written communication among the majority; it leaves opportunity for further learning to attain near 100% facility and it takes account of the educationally disadvantaged.

3) There was some evidence of Zipf's law (1949) operating in relation to word frequency and word length. It was found that other things being equal the mean number of strokes for characters of median difficulty was in the neighbourhood of fourteen strokes. Perhaps this is the magic number seven plus or minus two for Chinese orthography.

4) The study has raised a number of interesting problems and pointed to avenues of further research. One such is the need for basic research into the auditory reception of the language and the encoding of the symbols received. Another is the exploration of the cognitive aspect of grapheme-phoneme correspondence of the language. Chinese probably rates low in the requirements for good communication symbols. But in the balance and symmetry of its logographic form it probably carries more information per chunk and certainly has what Chao (1968) calls "elegance".

References

Burroughs, G.E.R. (1957) "A Study of the Vocabulary of Young Children" Oliver & Boyd

Chao, Y.R. (1968) "Language and Symbolic Systems" Cambridge University Press

Chuang, Chai-hsuan (1938) "A Fundamental Vocabulary of Chinese Characters" China Press (in Chinese)

Flood, W.E. (1957) "The Problem of Vocabulary in the Popularisation of Science" Oliver & Boyd

Kennedy, G.A. (1954) "The Minimum Vocabularies of Written Chinese" Yale University Press

Lenneberg, E.H. (Ed.) (1966) "New Directions in the Study of Language" M.I.T. Press

Miller, G.A. (1956) "The Magican Seven, Plus or Minus Two: Some Limits of our Capacity for Processing Information" Psych. Review Vol. 63 No. 2 pp. 81 - 97.

Rinsland, H.D. (1947) "A Basic Vocabulary of Elementary School Children" Macmillan, New York

Smith, F., Miller, G.A. (Ed.) (1966) "The Genesis of Language: A Psycho-linguistic Approach" MIT Press

Thorndike, E.L., Lorge I. (1944) "The Teachers' Word Book of 30000 Words" Columbia University Press

Watts, A.F. (1944) "The Language and Mental Development of Children" Harrap

Zipf. G.K. (1949) "Human Behaviour and the Principles of Least Effort" Addison-Wesley